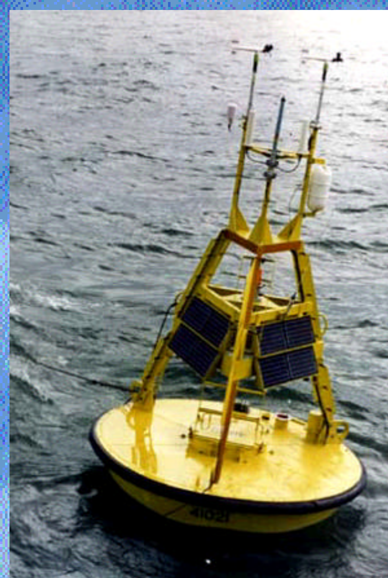
Assessing the Health of NOAA's Observing Networks

A Network Monitoring Improvement Plan for Determining Multi-Decadal Climate Variations



November 1998

Assessing the Health of NOAA's Observing Networks

A Network Monitoring Improvement Plan for Determining Multi-Decadal Climate Variations

Peter M. Steurer

U.S. DEPARTMENT OF COMMERCE

National Oceanic and Atmospheric Administration
D. James Baker, Administrator

National Environmental Satellite, Data, and Information Service
Robert S. Winokur, Assistant Administrator

Environmental Information Service
Susan F. Zevin, Deputy Assistant Administrator

National Climatic Data Center
Thomas Karl, Director

DRAFT

November 1998

On the Cover

The pictures on the front cover represent the diversity of NOAA's observing system networks that are used for operational purposes and also in measuring decadal to century scale climate variations. The picture in the upper left shows a station from the National Weather Service's (NWS) Cooperative Observer Program. Continuing clockwise, is a three-meter Olympic Buoy operated by the National Data Buoy Center. The NOAA South Pole Observatory is shown in the next photograph.This station is one of the baseline stations performing continuous trace gas measurements and is run by the Climate Monitoring and Diagnostics Lab. An upper air balloon is being readied for a launch in the next picture. This station is part of the NWS Upper Air network. An example of a ship from the Surface Marine Observing network is depicted next. This particular ship is operated by the NOAA Corps. The last picture is an Expendable Bathythermograph (XBT) launcher that is used to measure sea temperature profiles. The National Oceanographic Data Center collects data for this network.The photographs used on the cover and throughout this document were obtained from various NOAA web sites.

Acknowledgments

Many individuals contributed to the completion of this report. Claude Duchon from the University of Oklahoma and Matt Menne from the National Climatic Data Center (NCDC) provided significant input especially in the area of data homogeneity. The author would also like to thank many individuals from NCDC who gathered the data and produced the various indicators used to measure network performance. These include David Bowman, Charles Phillips, Peter Jones, Dave Smith, and Steve DelGreco. The design and layout of the cover was done by Scott Miller. Many others provided constructive comments and suggestions, especially NCDC Director Tom Karl and Global Climate Lab chief Rob Quayle.

This project was supported by the NOAA Administrator's Discretionary Fund. The effort was recommended by the NOAA Long-Term Climate Monitoring Council.

TABLE OF CONTENTS

Executive Summary

There is not now a structured process to regularly monitor the performance of NOAA's observing networks and report on the suitability of these networks in measuring decadal to century scale climate variations. This network monitoring improvement plan has been developed to routinely assess the adequacies of NOAA's observing networks in terms of their long-term climate monitoring capabilities.

NOAA operates many observing networks, usually for operational purposes. Some networks are considered important to climate change research and directly relevant to a NOAA Strategic Plan goal which is to Predict and Assess Decadal to Centennial Change. Ten networks that support NOAA's ability to achieve this goal and possibly other NOAA Strategic goals have been chosen as candidates for network health assessments. They are the Cooperative Observer Network, the U.S. Historical Climate Network, the Hourly and 15 Minute Precipitation Data Networks, the U.S. Surface Hourly Network - ASOS, the U.S. CLIMAT Network, the U.S. Upper Air Network, the U.S. Buoy Network, the Surface Marine Network, the Trace Gas (O_3 and CO_2) Networks, and the Ocean Profile Networks. Other NOAA observing networks, such as those that produce remotely sensed observations, may be addressed in future phases of this effort. Non-NOAA observing networks important to long-term climate monitoring also may be included at a later time.

In order to conduct an assessment of NOAA's observing networks in terms of their climate monitoring capabilities, quantitative performance indicators have been defined. These performance indicators are based upon the integration of common network characteristics of data and metadata with principal network observing stages such as those involved with a source-to-user life cycle of data. Adding another factor based upon the importance and practicality of performance measurement, results in a group of performance indicators that are viewed as most significant to decadal to century scale climate variations. These performance indicators are the number and spatial coverage of observing platforms in the data archive, the archive quality and completeness of the data, the length of the archived data record, the maintenance of the archived data homogeneity, and the data receipt timeliness. The indicators attempt to quantitatively address all aspects of network performance and are considered to be generic to observing networks. However, the specific definition assigned to each performance indicator can vary between networks and is dependent upon the parameter being measured, the method of measurement, and the intended application of the resulting data.

A network health assessment of the Cooperative Observer Network has been completed and is presented as a chapter in this report with an overall network assessment at the end of that chapter. As they are completed, assessments for other observing networks will be added as future chapters to this report. The health assessments presented in this report represent the performance indicators produced in a hindcast mode using the entire historical data base. It is expected that the hindcast mode will develop the historical base that would allow a more informed analysis of the current state of the networks. Regular assessments also will be performed in a near-real time monitoring mode and will take the latest data from the network and place it in a historical context. This latter mode will also produce near-real time assessments that identify problematic observing platforms. A password-protected World Wide Web prototype system that provides near-real time network assessments for the Cooperative Observer Network has been developed.

The hindcast and the near-real time monitoring health assessment reports attempt to characterize the strengths and weaknesses of NOAA's observing networks and will be provided to the responsible NOAA Network Managers. Observing network deficiencies will be clearly identified so that the requirements for corrective action can be implemented at the earliest possible time. It is hoped that any NOAA resources necessary to improve the observing networks are given consideration.

Methodology to Assess the Health of NOAA's Observing Networks
Chapter 1

1.1 Overview

The ability to accurately measure long-term climate variations can only occur if there is an ongoing program to regularly monitor the performance of the observing system networks, and if the results of the monitoring are taken into consideration by the responsible network managers (WMO, 1992). NOAA currently lacks a structured process to routinely monitor the performance of its observing networks and report on the suitability of these networks in measuring decadal to century scale climate variations. A network monitoring improvement plan has been developed which routinely assesses the adequacies of NOAA's observing networks in terms of their long-term climate monitoring capabilities.

NOAA operates many observing system networks with purposes ranging from predicting daily weather conditions, providing warnings for severe weather, managing water resources, aiding commerce by air and water, and understanding ocean dynamics. The majority of NOAA's observing systems exist for operational purposes. In view of these needs, most of NOAA's observing systems are managed according to their intended operational purposes and not for the after-the-fact use of understanding and predicting multi-decadal climate variations. The observing systems necessary to measure decadal to century scale climate change may be different from those that exist for operational purposes. Global climate change research requires systematic, long term, homogeneous, and accurate observations with sufficient spatial and temporal densities to monitor changes and variations of the climate system (IPCC, 1990). However, it is not known whether NOAA's observing systems are capable of providing the data necessary to fully understand climate change.

1.2 NOAA Observing Networks

Many observing networks exist throughout the NOAA organization that are important to its mission of environmental assessment, prediction, and stewardship. Some of these observing networks are also critical to one of NOAA's

ASOS Instrumentation at Miami Opa Locka, Florida

NOAA Observing Networks
Critical for Climate Monitoring

Cooperative Observer Network
U.S. Historical Climate Network
Hourly&15 Minute Precipitation Networks
U.S. Surface Hourly Network - ASOS
U.S. CLIMAT Network
U.S. Upper Air Network
U.S. Buoy Network
Surface Marine Network
Trace Gas (O_3 & CO_2) Networks
Ocean Profiles (salinity, temperature)

Figure 1

Strategic Plan goals which is to Predict and Assess Decadal to Centennial Change (NOAA, 1995). This goal states that "NOAA will provide science-based options for decisions regarding decadal-to-centennial changes in the global environment, specifically for: climate change and greenhouse warming; ozone layer depletion; and air quality improvement." Achievement of this goal requires improvements to the long-term climate record by enhancing domestic and international weather networks, observing procedures, and information management systems.

Figure 1 lists ten NOAA observing networks that have been chosen as candidates for health assessments. It is thought that these networks directly support NOAA's ability to successfully predict and assess decadal to centennial change as well as other NOAA goals such as seasonal to interannual climate forecasts. Future phases of this effort may address the health of other NOAA observing networks such as remotely sensed observations. Non-NOAA observing networks that are viewed as to be important to long-term climate monitoring also may be included at a later time.

1.3 Network Assessment Approach

In order to achieve and maintain a high level of efficiency and effectiveness within observing networks, an ongoing program of performance monitoring should be a fundamental part of each observing system (WMO, 1992). A method to conduct such an assessment in terms of NOAA's climate monitoring capabilities is to identify the critical components of NOAA's observing systems. It has been found that these critical components can be expressed in terms of three attributes consisting of common characteristics of data and metadata, principal observing stages, and the relevance of performance monitoring at the intersection. Combining these three factors yields a series of performance indicators that are important to climate change and will aid in identifying the strengths and weakness of an observing network.

A review of the literature has revealed that once an observing network is in place and operational, there are many characteristics of data and metadata that are common among observing networks (WMO, 1983; WCP-85, 1985; WCP-116, 1986). These common characteristics are listed in Figure 2 in an unprioritized list. They have been intentionally kept simple with the expectation that they can be applied to all NOAA observing networks. Thus, the data and metadata from all observing networks require some degree of quality

NOAA Observing Networks

Common Data & Metadata Characteristics

Quality
Timeliness
Completeness
Length of Data Record
Maintaining Homogeneity
Number of Observing Platforms
Spatial Coverage of Observing Platforms

Figure 2

with tolerable limits for missing information as measured by completeness. All required data and metadata should be received on time. The length of the data record should be sufficient for the intended analysis. Homogeneity of the data should be maintained and should be documented as metadata when it is not maintained. All observing networks require a certain number of observing platforms in order to provide adequate spatial coverage for their intended applications.

In order to transform these common data and metadata characteristics into facets applicable to each of NOAA's observing networks, specific definitions must be assigned that reflect the parameter being measured, the method of measurement, and the intended application of the data. For example, to effectively measure long-term climate variations, analyses may show that the spatial arrangements of Trace Gas observing platforms require one platform every 10,000 kilometers. However, an observing platform measuring daily maximum and minimum temperatures may be needed every 100 kilometers. Thus, the common characteristic *Spatial Coverage of Observing Platforms* is universal to these two observing networks. However, because of network diversity, the criteria defining adequate performance will vary.

In addition to the common characteristics of data and metadata listed in Figure 2, a second dimension to a comprehensive monitoring system can be added that describes all principal observing stages within a network. These principal stages outline the end-to-end management of data. They include the initial design of the network, sensor design and type, sensor density and siting, the actual observation, collection, transmission, receipt, processing, archival, access, and products supplied to users. These stages have also been referred to as a source-to-user life cycle of data.

By integrating the functionality of a network's principal observing stages with common characteristics of data and metadata, a group of performance indicators is produced that can be quantified. The outcome of the integration is shown in Table 1. The resulting matrix of performance indicators can be used to monitor observing system performance from an observation's inception to the creation of final products from those observations.

Another factor is also implied by Table 1 which addresses the relevance of the performance indicator and determines whether or not monitoring is actually needed at that level. The inclusion of an 'X' or 'x' in a matrix cell signifies that a performance indicator is thought to be relevant. Blank cells indicate performance indicators that are not considered directly relevant. For example, the common characteristic, *Quality*, is considered relevant across all principal observing stages and thus could have a performance indicator at each stage. The common characteristic, *Number of Observing Platforms* is

NOAA Observing Networks
Critical Components

Common Characteristics of Data and Metadata / *Principal Observing Stages*	*Quality*	*Timeliness*	*Completeness*	*Length of Data Record*	*Maintaining Homogeneity*	*Number of Observing Platforms*	*Spatial Coverage Observing Platforms*
Network Design	x	x	x	x	x	x	x
Sensor Design/Type	x	x	x		x		
Sensor Density/Siting	x			x	x	x	x
Actual Observation	x	x	x		x		
Collection	x	x	x				
Transmission	x	x	x				
Receipt	x	**X**	x				
Processing	x	x			x		
Archive	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Access	x	x	x				
Products	x	x	x	x	x	x	x

Network Performance for Climate Change
X = direct monitoring levels   *x = indirect monitoring levels*

Table 1

thought to apply during *Network Design*, *Sensor Density/Siting*, *Archive*, and in developing *Products*.

Relevant performance indicators are differentiated by ‘X’ or ‘x’ and have a degree of relevance associated with them. This relevance is largely dependent on the intended application of the data and the practicality of performance measurement. Thus, the performance indicators that have been deemed as most significant to climate change research have been highlighted as a bold red uppercase ‘X’. For example, *Quality* during the *Collection, Transmission*, and *Receipt* of data may be very important to a near-real time application such as providing a short-term forecast or warning. However, for assessing multi-decadal climate variations, the *Quality* of the final *Archive* may be considered of greater importance since additional time is usually available to improve the quality of data. Additional time is not always available in a near-real time application and thus *Archive Quality* is of lesser importance.

Some performance indicators listed in Table 1 are not always practical to measure retrospectively. For example, the *Completeness* of data that is expected when choosing a *Sensor Design/Type* should be decided when networks are first planned or when existing networks require more modern sensors. Thus, the design or selection of the appropriate sensor is a primary issue during the planning stage of a network. Once the *Sensor Design/Type* is chosen, the resulting *Completeness* of data is better assessed at a later observing stage more associated with the actual measurements from the sensor, such as the data *Receipt* from the sensor or the data that become the final *Archive*. In addition, monitoring at one principal observing stage may accentuate a network weakness at another observing stage. Thus, a data problem identified in the *Archive* may be a weakness associated with inadequate planning when the *Sensor Design/Type* was selected.

For each NOAA observing network, direct performance monitoring will occur at those levels in Table 1 that have been highlighted as a bold red uppercase ‘X’. The functions indicated by matrix cells that are a lowercase ‘x’ will not be directly assessed at this time. Note that almost all performance monitoring will be conducted at the A*rchive* level. For applications that involve decadal to century scale climate variations, the final *Archive* is considered one of the most critical components and also the most practical to measure retrospectively. As previously stated, an added benefit of monitoring the final *Archive* is that this level indirectly assesses the strengths and weaknesses of all observing stages (i.e., *Network Design, Sensor Design, etc.*). One exception is *Timeliness*. It is viewed that the *Timeliness* of the data and metadata can be effectively measured at either the *Archive* or *Receipt* level stages or at both stages.

1.4 Network Monitoring Improvement Plan

The highlighted performance indicators shown in Table 1 are the core ingredients of the plan that will be used to routinely assess the adequacies of NOAA’s observing networks in terms of their long-term climate monitoring capabilities. These performance indicators will be expanded into specific definitions tailored to each of NOAA’s observing networks as described above. Using these specific definitions, assessments will be conducted on each NOAA observing network show in Figure 1 and in the priority order presented in that list.

Assessments will be performed separately on each network and in two distinct modes of operation. These include a hindcast mode and a near-real time monitoring mode. The hindcast mode will use the entire historical data file and will be developed first. In this mode, the historical data base will be processed in order to provide an assessment of the magnitude of changes to the network over time. At either monthly or quarterly intervals depending upon the observing network, the near-real time monitoring mode will take the latest data from each network and place it in historical context. The historical mode will be run once and the near-real time monitoring mode will be run at regular intervals. Separate reports will be produced that describe both modes and will be provided to the responsible Network Managers. The contents of these two report types are more fully described in Section 1.5 along with their intended applications. The data from each network used to derive the performance indicators will be stored in a relational data base management system (RDBMS) that has the table resolution of an observing platform, year, month, measured parameter. These platform-level counts can then be aggregated into health assessment products for each observing network. For overall reporting purposes, the indicators will be produced at the network level but the capability also exists to track at finer levels. For example, surface land stations from the Cooperative Observer Network (COOP) can be aggregated into assessment reports by U.S. states and counties, National Weather Service (NWS) Weather Forecast Offices (WFO), and climate divisions. This capability allows a health assessment report to be issued for any performance indicator for the stations in the entire COOP network, or just for a given state and county, or just for the COOP stations that are the responsibility of the WFO in that state.

Maintaining the basic health assessment data in RDBMS allows an enhanced flexibility for tailored climate network health assessment products via Structured Query Language (SQL) statements. An example of a tailored product would be to list all COOP stations in California that were open for 100 or more years having less than 2 percent missing data and less than 1 percent of maximum temperature data flagged with quality issues.

1.5 Network Monitoring Reports

Two separate reports will be prepared that provide an assessment of the adequacies of NOAA's observing networks in terms of their long-term climate monitoring capabilities. The health assessments presented in these reports will represent network performance indicators produced in both a hindcast and a near-real time monitoring mode. Both reports will be provided to the Network Managers in the form of feedback to alert of problem areas and problem observing platforms within the networks.

As discussed above, the first report will be run in

COOPERATIVE OBSERVER MONITORING REPORT
DATA QUALITY - JULY 1998

This report identifies stations in the Cooperative Observer Network which have been determined to contain possible data quality inconsistencies. The following list of stations shows the total number days within the report month where inconsistencies were found.

State-Id Number	*Station Name*	*Temperature Inconsistencies Maximum*	*Temperature Inconsistencies Minimum*	*Precipitation Inconsistencies Rain*	*Precipitation Inconsistencies Snow*
AZ-7281-04	Roosevelt 1 WNW	31	0	0	0
AZ-6250-05	Parker	5	0	0	0
AZ-1026-06	Buckeye	4	0	0	0
AZ-0808-07	Black River Pumps	2	0	0	0
AR-1442-06	Claredon	31	0	0	0
AR-7712-06	West Memphis	0	31	0	0
AR-6768-08	Sparkman	1	31	0	0
CA-8163-01	Shelter Cove Av	1	0	0	0
CA-9440-01	Warm Springs Dam	31	30	0	0
CA-5378-04	Martinez Water Plant	31	0	0	0
CA-5866-04	Morro Bay Fire Dept	31	0	0	0
CA-8380-05	South Entr Yosemite NP	31	0	0	0
CO-6765-01	Parker Reservoir	1	0	0	0
CO-7309-01	Ruxton Park	31	31	0	0
CO-5414-02	Marvine Ranch	1	31	0	0
CO-6203-02	Ouray	0	31	0	0

Prototype WWW Near Real Time Monitoring Report for COOP Network
Figure 3

a hindcast mode using the entire historical data base. The intention of this report is to provide a historical assessment for each of NOAA's observing networks that would allow a more informed analysis of the current state of the network. These historical reports are included in this document. As the historical processing for each NOAA observing network is completed, the resulting assessment will be added as a separate chapter to this document. The COOP network has been complete and is included as Chapter 2.

The second report will be prepared at monthly or quarterly intervals and provide a near-real time monitoring assessment of each NOAA observing network. Information in this report will be reported at the observing platform level. Near-real time assessments at the observing platform level are anticipated to be one of the greatest benefits of network monitoring. The intention for near-real time assessments is to initiate corrective action for problematic observing platforms at the earliest possible time. This action would minimize the need for subsequent data archeology and data homogeneity efforts, which reconstruct the past climate after the fact.

At this time, network monitoring reports will be produced in hard copy form. World Wide Web (WWW) access to these reports is expected in 1999 from the National Climatic Data Center's (NCDC) Homepage. The capabilities of the WWW would allow rapid dissemination of information so that plans for corrective action could be implemented at the earliest possible time. A prototype WWW version of near-real time network assessments is available for the COOP network. The system, called Web CliServ, allows one to search station history files and learn of the existence of data and station history information. By clicking on "Additional *Options* available to NWS Offices only," one can preview a prototype network monitoring system that tracks performance indicators associated with data quality and receipt timeliness of COOP data. The system is flexible in that it produces assessment summaries for the entire network or for NWS offices which have responsibility for their COOP stations. A sample page from an assessment report from this system is shown in Figure 3. A password is required to enter the prototype monitoring system that can be acquired from the author of this report. The WWW address for Web CliServ is:
www.ncdc.noaa.gov/ol/climate/stationlocator.html

1.6 Limitations and Expectations

The network monitoring improvement plan outlined above, attempts to build an ongoing program of performance monitoring for NOAA's observing networks. As with all plans, there are limitations in areas that the plan does are not addressed as well as expectations for the desired achievements.

Limitations: To produce the information necessary for accurate and timely measurements and products, a comprehensive monitoring system ensures that all critical components of the observing networks, as listed in Table 1, are functioning efficiently and effectively. The network monitoring improvement plan defined above does not directly address those issues associated with the initial design of networks, whether the instrument sensors are of the proper design and are appropriate for the intended applications, and whether the products supplied to users are appropriate. Because these unaddressed issues are considered to be of paramount importance, a more comprehensive strategy that directly includes them may be necessary.

Expectations: The network monitoring improvement plan defined above produces a historical assessment of NOAA's observing networks. It also produces near-real time regular assessments. These assessment reports will be provided to the responsible Network Managers to alert them of problem areas and problem observing platforms within the networks. It is hoped that these reports are taken into consideration by Network Managers so that a plan for corrective action can be put in place at the earliest possible time. It is also hoped that the NOAA resources necessary to accomplish this are given consideration.

Assessment of the Cooperative Observer Network
Chapter 2

2.1 Historical Background of Network

The Cooperative Observer Network (COOP) consists of thousands of volunteer citizens and institutions throughout the United States that have systematically recorded weather information for over a century. These observers have provided and continue to provide basic weather information for their location, usually on a daily basis. Historically, about 32,000 observing sites have been part of the COOP network. In 1998 there were almost 12,000 sites that were actively participating in the network. The spatial distribution of these stations is shown in Figure 4.

The COOP network can trace its roots back to Colonial days (NRC, 1998). Thomas Jefferson envisioned a volunteer weather observer network in 1776. The Smithsonian Institute established a network of volunteer observers in 1847. A volunteer observer network was established by the Army Signal Corps in 1874 followed by the establishment of State Weather networks in the 1880s. The COOP network was officially established on October 1, 1890 with the enactment of the Organic Act. By this Act, all weather functions in the United States were transferred to a new agency called the Weather Bureau. One of the primary mandates was the volunteer observing program.

The most common COOP observations consist of once-a-day recordings of the maximum and minimum temperatures for the previous 24 hours, the temperature at observation time, 24-hour precipitation totals, 24-hour snowfall totals, and depth of snow at observation time. Some stations also supplement their observations with the daily weather occurrences, such as fog, thunder, hail, damaging wind, and the times that precipitation occurred. These are called "A" stations and are members of the Climate Network of the Cooperative Observer Program of which there are currently about 5,500 active stations.

A large number of stations measure only precipitation and / or river stages. These are called "B" stations and are members of the Hydrology Network. Another type, called "C" stations, is equipped with additional instrumentation and make more specialized observations such as soil temperatures and evaporation. The same station may be a member of the "A", "B", and "C" networks or just one network.

Cooperative Observer Climate Network
Observing Platforms Active in 1998

Figure 4

From the large network of COOP stations, a subset of approximately 1,200 stations was selected to form the United States Historical Climate Network (USHCN). Only those sites that were not believed to be substantially influenced by artificial changes of the local environment were included in the USHCN network (Karl, et. al., 1990). These stations also have a long period of record and tend to be located in areas with low population density.

2.1.1 Observing Equipment and Methods

COOP stations are equipped with standard instruments that are expected to meet observing requirements as defined by NWS. The equipment may be owned by the NWS, the observer, or a company or other government agency. Observations are made once a day either in the early morning, late afternoon, or midnight. The observations span the weather conditions over the previous 24 hours. Typical instrumentation at a COOP station is shown in Figure 5.

Most COOP stations are equipped with a Standard 8-inch Rain Gage (SRG) to manually measure precipitation. An observer makes a precipitation observation once a day by placing a dip stick inside the SRG and reading the amount. Snowfall and snow depth are also measured manually each day with snow stakes or rulers. About 3,000 sites have recording gages that measure precipitation as a function of weight and 15 minute or hourly time intervals. These stations are part of the Hourly and 15 Minute Precipitation Networks described in a later chapter.

Many COOP sites measure temperature and use either a Cotton Region Shelter (CRS) or a Maximum / Minimum Thermometer System (MMTS). The CRS contains liquid-in-glass maximum and minimum thermometers that are used to measure the daily extremes of temperature. The MMTS uses a thermistor that electronically saves the maximum and minimum temperature for the previous 24 hours.

Some stations make other meteorological or hydrological observations. A small number have evaporation pans, soil temperature instruments, and make other measurements related to agriculture. Many stations have river gages to support NWS hydrology and warning programs.

2.1.2 Data Communication and Processing

The majority of all daily observations made at COOP stations are recorded manually and transcribed on monthly paper forms by the observer. The forms are retained at the station for a calendar month. Some stations also transmit their data near-real time to the NWS. The near-real time mode and manual mode of data transmission are essentially two separate paths that do not meet. The manuscript forms are digitized and have a high level of Quality Control (QC) performed. The data eventually end up in the official archives after many months. Data that are received daily in a near real time mode are used by the NWS to assist in forecasts and warnings. Observers transmit the data by telephones either as direct voice communication or keypad entry of the data. This form of transmission has been found to be less reliable than manually receiving forms (NRC, 1998). In addition, all observations are not transmitted, such as significant weather events. Even though the manual mode of transmission is much slower and more labor intensive, it is currently the preferred form of data transmission and is described more fully below.

At the end of a month, the observer mails the manuscript forms containing that month's data to the responsible NWS forecast office. NWS performs a preliminary quality review of the forms, checking for obvious errors. The forms are then forwarded to NCDC where the manuscript records are digitized and a high level of manual and automated QC is performed. NCDC publishes the data monthly in a hard copy periodical called Climatological Data. The original forms are maintained at NCDC, either on microfiche or as scanned digital images. The keypunched digital data are archived on magnetic media with a primary copy and an off-site back up. The digital file retains the original observed values and any edited values that may have been created during the NCDC QC. If all goes well, daily observations will be available for analysis 90 days after the end of a data month (ie. All observation made between January 1-31 would be available for analysis after the end of April).

NCDC also maintains extensive metadata concerning the COOP network. One type of the metadata is the data set documentation which describes the record layout, historical changes to the network, known problems, possible biases in the data, and many other aspects important to data users.

At the station level, NCDC maintains an extensive station history data base that tracks metadata characteristics associated with a station. The metadata include changes in location, elevation, instrumentation, observation time, exposure, observer, as well as other parameters. Station history forms, called B-44s, are prepared by NWS officials when changes occurs at a station and are then mailed to NCDC. The information on the B-44 form is QCed and then entered into a relational data base. Examples of common changes are station relocations, observer changes, and instrument replacement. As with the data, many months (or even years) can occur between the time a B-44 form is received and the time the actual change occurs at a station.

Even small changes in site characteristics at a station can introduce biases into the historical observation record. For climate change studies, it is important to document changes to a COOP site over time so that users can adjust for artificial changes that have affected the historical record. (Karl, et.al., 1990).

St. Johnsbury, Vermont
A participant in the
Cooperative Observer
Weather Network since 1894

A Cotton Region Shelter containing liquid-in-glass maximum and minimum thermometers is shown at the right. Precipitation is measured by a standard 8 inch rain gage shown at the left and a Fischer-Porter recording rain gage located in the center. Snow stakes are to the left along the fence with a whisk broom used to sweep the snow board.

Figure 5

2.1.3 Network Responsibility

Within NOAA, NWS and NCDC are responsible for managing the COOP network. Generally, NWS is responsible for all processes that involve the actual collection and transmission of observations. NCDC is responsible for procedures that involve the long-term archival and dissemination of the observations.

The NWS manages the COOP network under the auspices of the Cooperative Observing Program. The NWS responsibilities for network management are large and diverse. They range from selecting the appropriate observing site; recruiting, appointing, paying, and training observers; installing and maintaining equipment; documenting the station history; data collection and delivery to NCDC; performing an initial quality review; and maintaining overall fiscal and human resource responsibility. Each NWS forecast office is assigned the responsibility of managing COOP stations within their geographic area. At present, Data Acquisition Program Managers (DAPMs) are staffed at each WFO and perform COOP network management duties as part of their many other responsibilities. Management oversight of the DAPMs is maintained by Regional and National Cooperative Program Managers and

COOP Network Performance Indicators

Performance Indicator	Definition
Archive / Number of Observing Platforms	Number of stations available
Archive / Spatial Coverage Observing Platforms	Gridboxes (or counties) with stations
Archive / Length of Data Record	Number of stations open 100+, 50+, 25+, 10+ Years
Archive / Completeness	
Temperature (Maximum & Minimum)	Percent of data present
Precipitation	“
Snow	"
Archive / Quality	
Temperature (Maximum & Minimum)	Percent of data with no quality issues
Precipitation	“ “
Snow	" “
Receipt / Timeliness	Percent of stations received on-time
Archive / Maintaining Homogeneity	Percent of stations no detectable homogeneity issues

Table 2

the WFO Meteorologist in Charge. NRC (1998) found that the priority assigned to the COOP program varies among the WFOs and that staffing may not be adequate to manage the COOP network.

After the observations have been collected by NWS, they are transmitted to NCDC. NCDC performs automated and manual QC on the data, provides dissemination of the data to users, and is responsible for the statutory requirements for final archival. NCDC also develops products from the data such as 30-year climate normals and assessments of the climate. It also performs periodic re-analysis of the data using contemporary data processing schemes.

2.2 Network Health Assessment Results

A network health assessment of the COOP network has been completed. Performance indicators have been defined in order to conduct a quantitative assessment of the strengths and weaknesses of the COOP network. The indicators were developed based upon the methodology described in section 1.3 of this document.

Seven performance indicators tailored to the COOP network have been determined as important factors in efforts to measure decadal to century scale climate variations. These indicators and their definitions are listed in Table 2.

The assessment results presented here represent the performance indicators being produced in a hindcast monitoring mode. The entire historical data base has been used which contains approximately 16 gigabytes of data. Observing platforms for the COOP network consist of fixed point sites called stations. Generally, the assessment reflects the data available for the years 1888 through 1998. Some stations also exist prior to 1888 but their numbers decrease rapidly prior to this time. Each performance indicator is described in a separate section below.

2.2.1 Archive / Number of Observing Platforms

Figure 6 depicts a time series of the number of stations that have existed in the COOP network each year from 1888 through 1998. Two general classes of stations are represented: those that observed both temperature and precipitation and those that only observe precipitation. As described above, these are called "A" and "B" type COOP stations, respectively. Distinction is also made between digital and non-digital stations. Digital is defined as stations that have been keypunched and reside in NCDC archives (NCDC, 1998). Non-digital is defined as stations that only reside in manuscript form in NCDC archives.

An analysis of Figure 6 indicates large and sometimes abrupt changes in the number of stations that constitute the COOP network. The upper line shows all stations that exist in both digital and non-digital form. Included in this group are non-digital stations that only observe river stages. An estimated 800 stations were "River only stations" in 1998. The lower two time series are a subset of the upper time series. They represent those stations that have digital temperature and precipitation data. Note that the number of stations that observe only precipitation is significantly greater than those that observe both temperature and precipitation.

NOAA Cooperative Observer Climate Network
All Stations - Entire Network
Performance Indicator: Number of Observing Platforms

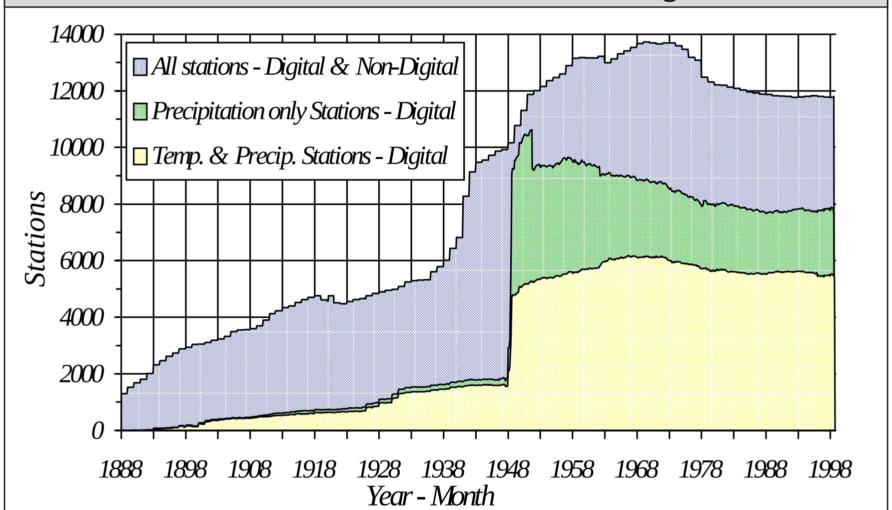


Figure 6

The upper time series shows a gradual increase in the number of stations the mid 1930s, followed by an exponential increase through the mid 1950s. This large increase is primarily attributed to an increase in stations measuring precipitation, established for hydrological projects such as dams and flood control. A smaller increase occurred from the 1950s through the early 1970s when the COOP network attained its maximum number of stations. During the remainder of the 1970s, a rapid decline in stations occurred which is attributed to the budgetary restructuring at the NWS. Since 1980, the network has been fairly stable but has exhibited a small decline in stations.

The two lower time series lines represent digital data and depict a somewhat similar pattern, except for an abrupt and very large increase in the number of stations occurring in 1948. This was the year that the Weather Bureau established Regional Weather Processing Units and the operational keying of COOP records began (Barger and Nyhan, 1960). In January 1952, the Regional Weather Processing Units were consolidated into one National Weather Records Center and located in Asheville, North Carolina (later to be renamed NCDC). A small but abrupt decrease in the number of digitized precipitation stations occurred at this time, likely due to this restructuring. Since 1952, the number of digitized precipitation stations has gradually declined whereas stations observing temperature have remained rather stable.

The arithmetic difference between the top two lines represents the number of stations that reside only in manuscript form. For example, in 1998 there are almost 4,000 stations that remain non-digital and in 1948 there are more than 8,000 undigitized stations. These data are potentially available for analysis but are currently very difficult to access. In general, many non-digital stations prior to the 1940s observed both temperature and precipitation while the stations since then observed only precipitation.

Digitization of COOP records prior to 1948 would vastly improve the completeness of the network. NOAA is currently performing a data rescue effort to digitize these pre 1948 COOP records. The effort has begun with the 1,200 USHCN stations and will be followed by other long term climate sites. A 5-year rescue effort to digitize the corollary station history records for all COOP sites prior to 1948 was completed by NCDC in 1998. The need for large scale digitization of COOP records after 1948 is more debatable since a large number of stations are already digital and many stations since 1948 observe only precipitation.

2.2.2 Archive / Spatial Coverage of the Observing Platforms

The spatial coverage of observing platforms provides a measure of whether all geographic areas within the observing network have a sufficient number of stations for the intended data application. The spatial coverage of COOP stations required for climate change analysis is considered to be beyond the scope of this report. However, station density guidelines do exist that can be measured. NWS (1953) prescribes that the COOP network contains one station every 25 miles or one station per 625 square miles (NRC, 1998). WMO prescribes that stations observing temperature exist every 50 kilometers and stations observing precipitation exist every 25 kilometers (WMO, 1983). For developing a reference climatological station network of monthly temperature and precipitation, a density of two to 10 stations per 250,000 square kilometers is considered necessary (WCP, 1986). NRC (1998) interviewed different users of COOP data and found that NWS indicated that at least one station should exist in every county in the United States to support forecast and warning programs.

Using the station density guidelines described above, two measures of spatial coverage for COOP stations have been assigned. These are a minimum of one station every 625 square miles and a minimum of at one station in every U.S. county. They are depicted in Figures 7 and 8 as gridded performance indicators. Figure 7 shows a time series of the percent of 25 mile grid boxes and county polygons that contain one or more stations for the period 1888 through 1998. Figure 8 is a map display of the 25 mile grid and county polygons at 50 year intervals for 1898, 1948, and 1998. All stations in the COOP network are included in these analyses as described in section 2.2.1. This would include stations not in digital form and stations that only observe precipitation.

NOAA Cooperative Observer Climate Network
All Stations - Entire Network
Performance Indicator: Spatial Coverage of Observing Platforms

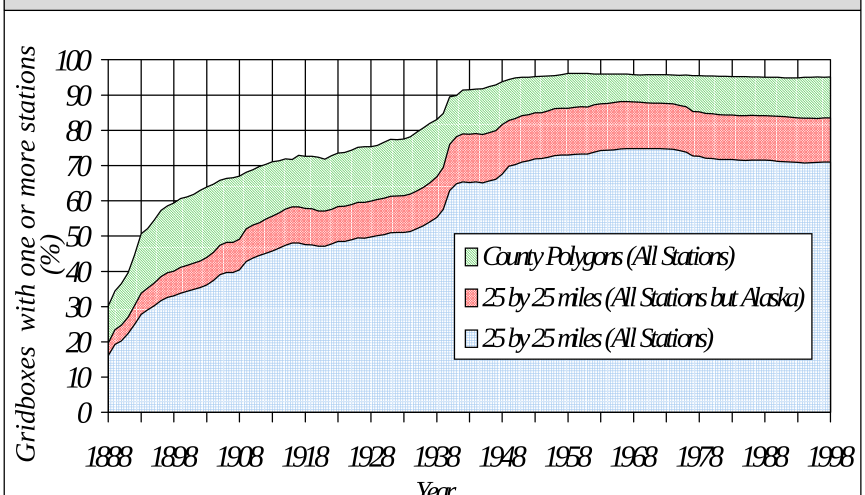


Figure 7

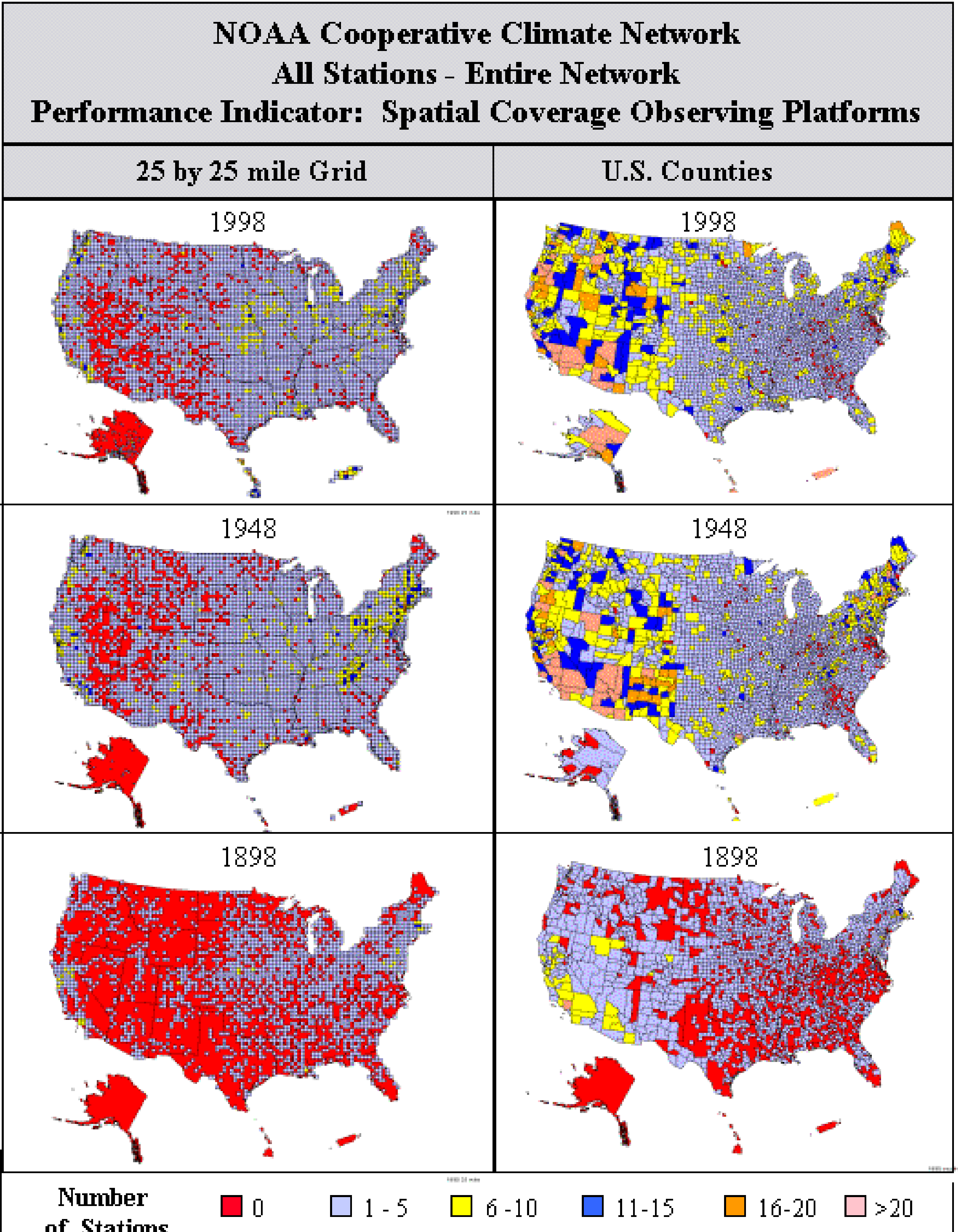

NOAA Cooperative Climate Network
All Stations - Entire Network
Performance Indicator: Spatial Coverage Observing Platforms
25 by 25 mile Grid
U.S. Counties
1998
1998
1948
1948
1898
1898
Number of Stations
0
1 - 5
6 -10
11-15
16-20
>20

Figure 8

An examination of Figures 7 and 8 reveals that more than 95% of the 3,140 counties in the United States contain at least one COOP station from the 1940s through 1998. Generally, counties without stations are geographically quite small and are located in the southeastern quadrant of the continental U.S. Georgia, eastern Virginia, and northern Kentucky contain the highest concentration of counties without stations. The western U.S. and Alaska almost always have at least one station per county after the 1940s since counties are geographically quite large.

The 25-mile grid presents an entirely different perspective of station distribution as compared to county polygons. Whereas the existence of stations within counties is most dependent upon geographic size of the county, population density appears to be the overriding factor here. Alaska and numerous areas in the desert and mountainous western region of the U.S. do not contain a station within a 625 square mile area. Maine, northern Minnesota, and localized pockets in the southern states also contain higher concentrations of grid boxes without stations. Since Alaska has relatively few 25 mile grid boxes that contain a station, Figure 7 depicts time series analyses that include and exclude Alaska. Thus, there are 6,193 total 25 by 25 mile grid boxes in the United States including Alaska. In 1998, more than 68 percent had at least one station. Excluding Alaska, there are 5,098 grid boxes and 82 percent had at least one station in 1998.

A detailed analysis of Figure 8 reveals that some overlap of 25 mile grid boxes and county polygons occurs where there is no station during 1998. These may be prime candidate areas to establish a new COOP station.

NOAA Cooperative Observer Climate Network
All Stations - Entire Network
Performance Indicator: Length of Data Record

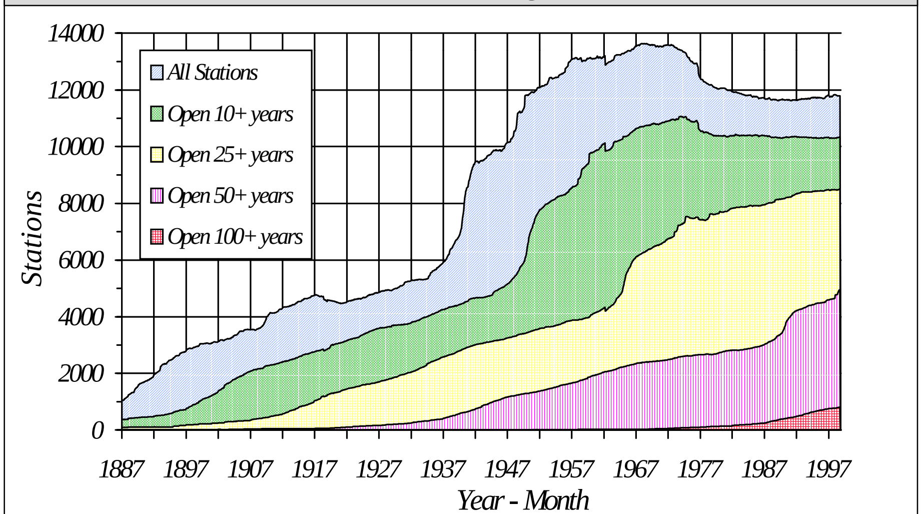


Figure 9

2.2.3 Archive / Length of Data Record

Figure 9 shows a time series of the number of stations that have been open for varying numbers of years from 1888 through 1998 and can be used as a performance indicator for length of data record. For comparison purposes, the total number of COOP stations that existed during every month of the 1888-1998 period is presented as the upper time series graph. The next graph from the top indicates the number of stations open for 10 or more years, followed by 25, 50, and 100 or more years, respectively. In October1998, for example, a total of 11,785 stations was part of the COOP network. Of this number, 10,349 stations had been open for 10 or more years, 8,519 for 25 or more years, 4,958 for 50 or more years, and 805 for 100 or more years. Stations were not included in these length of record totals if significant breaks occurred during that station's record. A significant break is defined as an interruption of 18 consecutive or 24 total months where observations were not carried out at a station (NCDC, 1988).

The time series graphs presented in Figure 9 represent all stations in the network. Since all stations are not in digital form as described above in section 2.2.1 and some stations only observe precipitation and not temperature, this is a more optimistic representation of the length of data record available in COOP network. However, even with non-digital and precipitation only stations included, the trend toward an increasing number of long-term stations is quite evident. An examination of the time series lines indicates upward inflection points occurring at 1947, 1962, and 1987 for the 10+, 25+, and 50+ time series lines, respectively. This is due to the large increase in stations that opened, beginning in 1937 and continued through the early 1970s. Assuming this pattern continues and stations remain open with the same consistency as in the past, the number of stations open for 100 or more years will gradually increase over the next several decades followed by a rapid increase beginning in the year 2037.

NOAA Cooperative Observer Climate Network
Digitized Stations - Entire Network
Performance Indicator: Archive Completeness

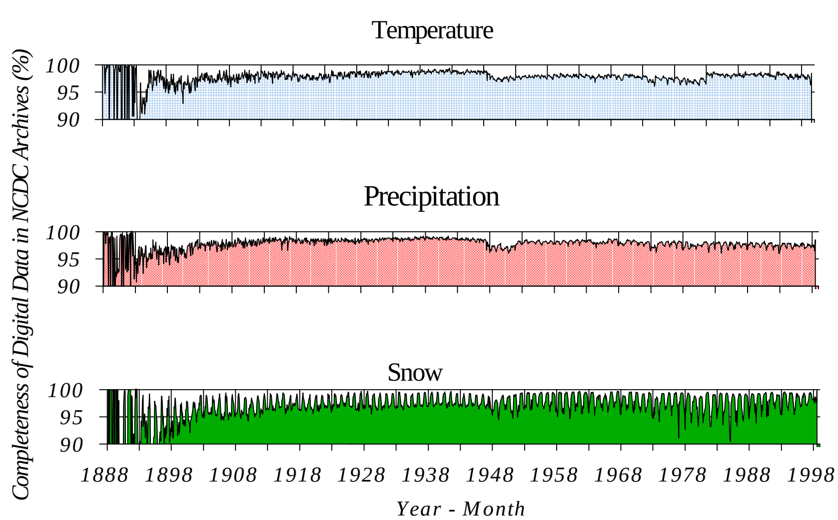


Figure 10

2.2.4 Archive Completeness

For the COOP network, completeness is measured as the amount of data in the final digital archive. Figure 10 shows completeness as a year-month time series of the percent of digital data available in the archive as compared to the total amount of possible digital data. For example, in January 1947 and January 1998, there were 1,831 and 7,770 stations, respectively, that have precipitation data in the COOP archive. Multiplying the number of days in the month by the number of stations results in the total possible days of precipitation for that month. For January 1947 and 1998, this would equate to 56,761 and 240,870 possible days, respectively. As shown in Figure 10, archive completeness for these two Januaries were both about 98% which means that about 2% of the total possible days were missing. This corresponds to 1,141 missing days for 1947 and about 4,929 missing days for 1998.

As defined here, archive completeness does not quantify the amount of potential data that may be available in non-digital form. This is measured in section 2.2.1 which assesses the Number of Observing Platforms.

The time series graphs subdivide the data into the main weather elements observed at a COOP station. These are temperature, precipitation, and snow. The percent of missing data for all elements ranges from 1 to 4 percent since about 1900. Prior to 1900 large variations in completeness occur. Some months have completeness percentages falling well below the 90 percent level and are not displayed on the graphs. These low percentages are not significant because only a small number of stations are in the digital archive prior to 1900 as shown in Figure 6. Therefore, a few stations with large amounts of missing data are substantially influencing the completeness percentage for the entire network.

A sinusoidal oscillation exists on the graphs, especially for snow, but is difficult to notice because of the resolution of the graphs. The peaks of these oscillations represent winter months and the valleys represent summer months. A plausible explanation is the increased difficulty in making a weather observation in winter as compared to summer. Supporting this assumption, the same oscillation is also evident in archive quality as described below.

2.2.5 Archive Quality

For the COOP network, the performance indicator for quality is a measure of the amount of data in the final digital archive that has passed all phases of NCDC quality control. NCDC quality control consists of operational processing and periodic rehabilitations of the historical archive.

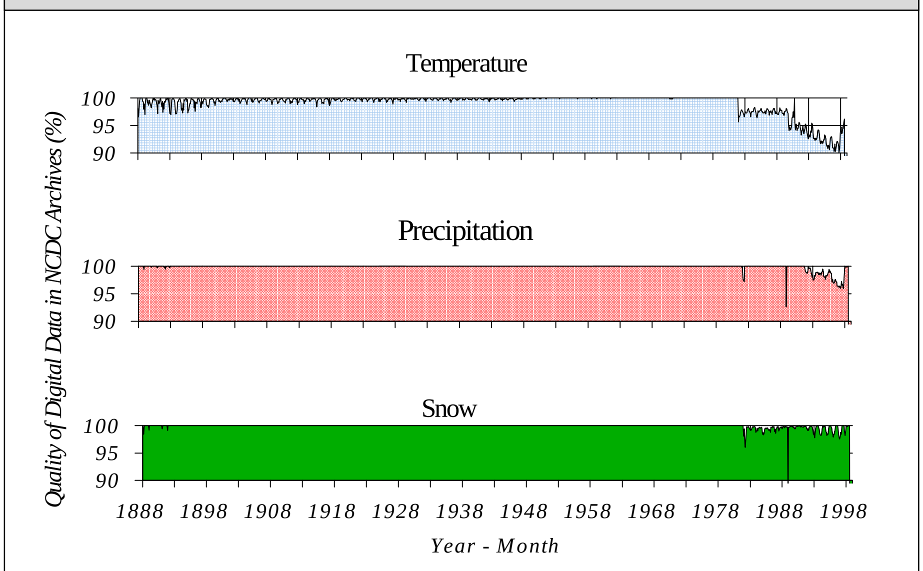


Figure 11

Figure 11 depicts archive quality as a year-month time series of the percent of unflagged data in regard to the total amount of data in the digital archive. Since 1982, a data value that operationally failed a quality test has been assigned a specific quality flag appended to the digital data (NCDC, 1998). In most instances, a replacement value is also included in the archive for these flagged data values. Prior to 1982, quality flags were not operationally assigned to data values. Since archiving and formatting practices were not accommodating to quality flags, all digital data values that were determined as erroneous prior to 1982 were replaced with an estimated value or, in some cases, deleted. No indications of these data changes were made in the digital archive. This practice is quite apparent in the Figure 11 time series graphs since most months prior to January 1982 have achieved close to a 100 percent quality level. Those months prior to January 1982 that have less than a 100 percent quality level are the result of historical rehabilitations of the data that took place in 1983 and 1995 (NCDC, 1998).

The time series graph subdivides the data into the main weather elements observed at a COOP station. These are temperature, precipitation, and snow. Since 1982, the percent of data that failed quality testing is between 3 to 8 percent for temperature and 4 percent or less for precipitation and snow. QC for these latter two elements is much more difficult and few quality issues are found as compared with temperature. Precipitation and snow QC mainly consists of limit checks whereas temperature also receives an areal QC with nearest neighbor stations. Data homogeneity work as described in section 2.2.7 and Appendix A intend to improve the QC capabilities for these three weather elements.

The sinusoidal oscillations also exist for the graphs in Figure 11 as described above under archive completeness. Again this oscillation is attributed to the difficulty in performing an observation in winter as compared to summer. In this case however, the difficulty results from an observation being made and flagged for quality. In the case of completeness, the difficulty results in an observation not being made at all.

A trend toward lower archive quality level is evident from 1988 through 1998. This was the period when NCDC put in place a new processing system for COOP data which allowed additional data errors to be found. As the system was enhanced through the 1990s, additional data were flagged as shown by the graphs. These lower quality levels in the 1990s are assumed to be more likely due to this enhanced capability of the QC system as opposed to a decrease in quality of COOP observations.

2.2.6 Receipt Timeliness

Receipt timeliness is a measure of how quickly an observation enters into data processing steps and becomes ultimately available for user applications. As described earlier in this chapter, COOP data are received in both manuscript and near real time modes. Since the manuscript mode is considered to have more accurate data and all COOP stations do not have the capability to transmit near-real time, the manuscript mode is the main method used for data receipt at NCDC.

A time series measuring COOP data receipt timeliness is presented in Figure 12 for the period July 1996 through October 1998. Prior to July 1996, consistent data receipt statistics were not maintained. This time series tracts only those stations that are digitized and currently includes about 8,000 COOP sites (see section 2.2.1 and figure 6). As an exception, note that the lower time series line represents about 250 Automated Surface Observation System (ASOS) stations that are also members of the COOP network. NCDC acquires these data in a near-real time mode and thus manuscript forms are not received for ASOS sites and not digitized.

COOP manuscript records contain an entire month of daily observations for a single station. The paper forms are sent to NCDC usually by regular mail. They are considered late if they are not received within 30 days after the end of the data month (i.e., A manuscript form containing June data needs to arrive at NCDC by July 30). Manuscript records arriving between 31 and 60 days after the end of the data month are considered late but are still processed but do become part of the digital archive. Any form received after 60

NOAA Cooperative Observer Climate Network
Digitized Stations - Entire Network
Performance Indicator: Receipt Timeliness

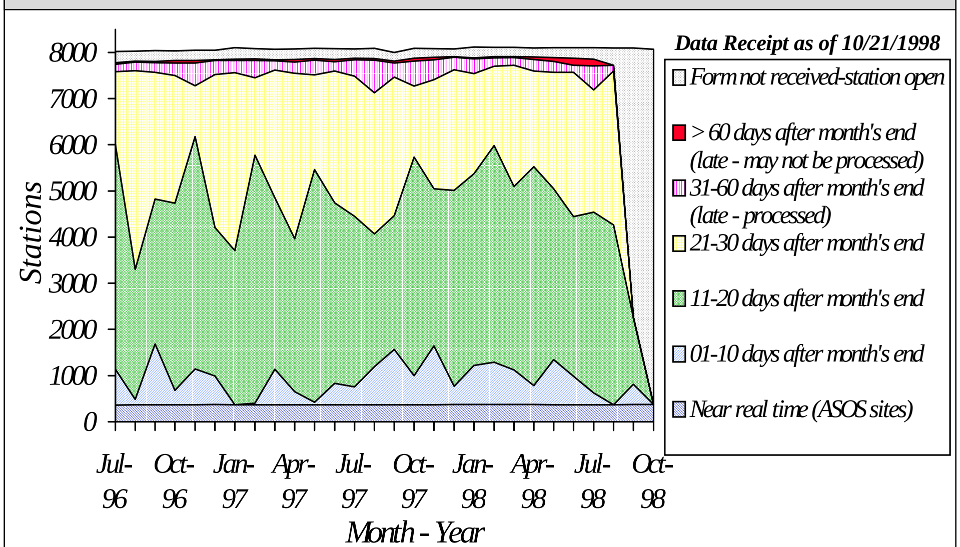


Figure 12

days is usually not processed and usually does not become part of the digital archive. These unprocessed forms are filed and become part of the non-digital archive. Some forms are never received due to observer trouble, mail problems, etc.

Figure 12 presents the cumulative number of stations received within specific day intervals after the end of a data month. In July 1996, for example, forms were received within 10 days after the end of the July data month for about 1,000 stations. A cumulative total of about 6,000 stations had been received within 20 days and about 7,600 were received within 30 days. Data receipt for these 7,600 stations were considered on time. Approximately 200 stations were received late (within 31 to 60 days), but were processed and did become part of the final digital archive. A small number of stations (10) sent in their forms after more than 60 days and approximately 200 stations never sent their forms. These latter two groups are the more problematic and are candidates for corrective action since the resulting data never enter the digital archive.

2.2.7 Maintaining Homogeneity

Data homogeneity assessment is a fundamental component of data quality. Homogeneity testing seeks to identify trends or discontinuities in the data record that result from non-climatic influences. Trends in the data may occur for reasons such as urban warming, instrument drift or other effects that accumulate gradually. Discontinuities may result from instrumentation changes, station moves, or possibly even lightning strikes, and result in sharp changes in the data record. Data homogeneity testing has generally been conducted retrospectively (e.g. Karl and Williams, 1987; Karl et al., 1988) and exploits the longer time series available in the data archive. In monitoring the climate networks, however, it is desirable to detect data homogeneity problems in near-real time so problems can be addressed as soon as possible. In this case, a much shorter data record is available for analysis and this imposes limitations on the magnitude of inhomogeneities that are detectable (see Appendix A). Nevertheless, work is proceeding with evaluating near-real time homogeneity testing using a type of "buddy" check, that is, comparing daily observations at one station to observations at surrounding stations.

The theoretical basis of this "buddy" check is that a random time series should result when the daily departures from the monthly mean temperature at one station are subtracted from the daily departures at a neighboring station. If the departure-difference time series is not a random series or if it is correlated with another departure-difference series derived using a different neighbor, then there may be some kind of data error present in the record at one of the stations. Work on evaluating this approach has revealed that highly correlated daily departure difference time series can result when errors are present in the data record of the station under consideration. The most common type of error detectable in this approach appears to be the "partial" shifter. A shifter refers to an observer who may attribute a temperature reading to its perceived day of occurrence, rather than recording the observation on the day it was observed. The qualifier "partial" is used since only a portion of the month is shifted. Current quality control procedures are very effective at detecting shifts that occur systematically throughout the month and this correlation analysis represents an enhancement of procedures already in place. It is also possible with this method to detect values at a station that are out of range with respect to its neighbors and, at least potentially, to detect errors of trend and step changes (true inhomogeneities). The limitations of near-real time inhomogeneity detection are discussed in Appendix A.

As an example of how near-real time inhomogeneity testing can be used, Figure 13a is presented showing the percent of stations with no detectable error or homogeneity issues. The analysis includes all COOP stations that observe temperature located in ten north-central U.S. states (Michigan, Indiana, Kentucky, Missouri, Illinois, Iowa, Wisconsin, Minnesota, North Dakota, South Dakota). Approximately one thousand stations were evaluated for errors and potential inhomogeneities during 1997 and through July

1998. Percentages for both maximum and minimum temperatures are presented. As Figure 13a indicates, in each month generally 5% or fewer stations were flagged with a potential error or inhomogeneity. Although each of the stations that make up the percentage was flagged with a potential data error, it does not mean that all data for that station during the month are questionable, only that at least one of the daily values resulted in highly correlated paired-difference time series and is probably in error. Thus, the percentage of data that may be in error is substantially less than the percentage of stations flagged with potential errors. Further, some minority of stations flagged with potential errors may have some legitimate weather-related explanation for the apparent error or there may simply be no obvious origin of the error and therefore no clear corrective measures.

Figure 13b indicates the number of months during 1997 that the same station was flagged. About half of the stations with potential data errors or inhomogeneities were flagged only once during 1997 and 44 stations in the case of daily maximum temperature and 27 in the case of minimum temperature were flagged for three months or more. The intent is that this near-real time homogeneity analysis will be an enhancement to the quality control procedures already in place and stations that are flagged repeatedly can be noted for corrective action.

NOAA Cooperative Observer Climate Network
Digitized Stations - North Central United States
Performance Indicator: Maintaining Homogeneity

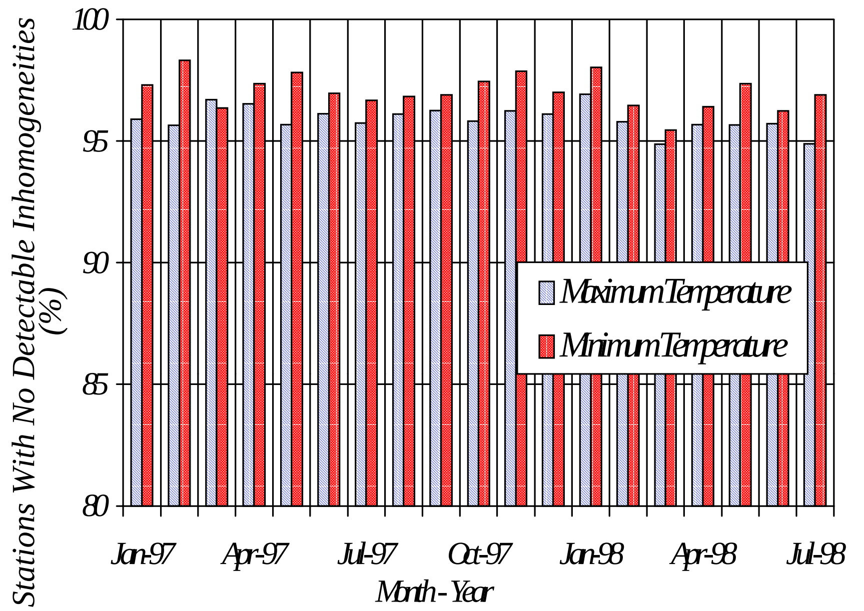


Figure 13a

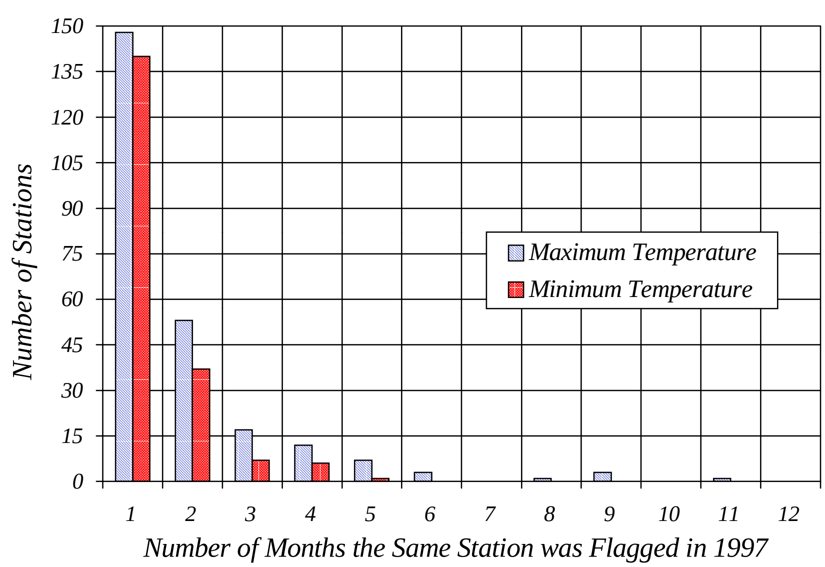


Figure 13b

2.3 Overall Network Assessment

A health assessment of the COOP network has been completed and is presented in this report. Other recent reports have indicated that the overall COOP network has critical deficiencies and may be at risk due the lack of spare parts for observing and processing sensors (NOAA, 1998). The NRC (1998) issued an authoritative report that cited many recommended improvements. The report presented here analyzes the COOP network in terms of its long-term climate monitoring capabilities. Quantitative performance indicators have revealed possible strengths and weaknesses in the network. An overall assessment of the COOP network in terms of each performance indicator is summarized below. Suggestions have been put forth based upon a qualitative analysis of the information presented in this report. It is suggested that this report be provided to the COOP Network Managers for analysis and comment.

Number of Observing Platforms: The COOP network currently contains 8,000 stations with digitized precipitation data and 5,500 stations with digitized temperature data. Prior to 1948, relatively few stations are in digital form. A project to digitize the pre-1948 data is underway. It is suggested that this data rescue work and the efforts necessary to enter the resulting keyed data into the archive remain a high NOAA priority.

Spatial Coverage of Observing Platforms: An analysis of the COOP network in terms of a 25 by 25 mile grid and U.S. county polygons indicates areas that do not contain a station. It is suggested that an analysis be conducted to determine those critical geographic areas where one or more stations should be established. The overlapping areas of 25 mile grid boxes and county polygons that do not contain stations are prime candidates for the establishment of new COOP stations.

Archive Length of Record: Due to the expansion of the COOP network from the 1930s through the early 1970s, the number of stations that will be open for 100 or more years is expected to increase well into the next century. This is considered a major strength of the COOP network. It is suggested that an analysis be conducted to identify those stations that will be entering the 100 or more years of record category. In addition to the length of record, the analysis should emphasis those issues important to climate change, such as site and observation stability. The identity of these long-term stations should be made known to COOP Network Managers so that efforts can be made to keep them open with minimal changes.

Archive Completeness and Quality: Between 5 and 10 percent of COOP data that are currently received each month are either missing or have quality issues. It is suggested that feedback be provided to COOP Network Managers to identify those stations that regularly contain missing data or have quality issues so that corrective action can be initiated.

Receipt Timeliness: Receipt of COOP data is slow and labor intensive due to the manual processing of paper forms. This is an obvious and major weakness of the COOP network. A modern data collection and transmission system is needed. The current manual process, however, appears to be reliable in terms of completeness and quality of data. Due to the importance of the COOP data, it is suggested that any redesign of the network be done with climate research as an important component. Prior to the implementation of a new data system, it is also suggested that COOP Network Managers be made aware of those stations that regularly send their data forms late or not at all so that corrective action can be initiated.

Data Homogeneity: The successful detection of inhomogeneities in the COOP record is a fundamental component of data quality. Early detection minimizes the need for subsequent data archeology and homogeneity efforts as well as produces a more accurate assessment of the current state of the climate. The exploration of statistical methods to detect inhomogeneities in the COOP record is currently in a formative stage. It is suggested that the NOAA resources needed to continue this work are given a high priority.

References

Barger and Nyhan, 1960: *Climatology at Work - Measurements, Methods, and Machines.* Washington, D.C., U.S. Dept. of Comm., 109 pp.

IPCC, (Intergovernmental Panel on Climate Change), 1990: *Climate Change: The IPCC Scientific Assessment*, J.T. Houghton, G.J. Jenkins and J.J. Ephraums (eds.). Cambridge University Press, Cambridge, UK, 365 pp.

Karl, T.R., and C.N. Williams Jr., 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Meteor.*, **26**, 1744-1763.

----- , H.F. Diaz, and G. Kukla, 1988: Urbanization: Its detection and effect in the United States climate record. *J. Climate*, **1**, 1099-1123.

-----, C.N. Williams, Jr., and F.T. Quinlan, 1990: *United States Historical Climate Network serial temperature and precipitation data.* ORNL/CDIAC-30, NDP-019R1. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 374 pp.

NCDC, 1988: *Environmental Information Summaries C-27: 1951-1980 Climatic Normals.* Asheville, N.C., U.S. Dept. of Comm., 9pp.

NCDC, 1998: *Surface Land Daily Cooperative Summary of the Data - TD3200: NCDC Reference Manual Series*. Asheville, N.C., U.S. Dept. of Comm., 33pp.

NOAA, 1995: *NOAA Strategic Plan: A vision for 2005: Executive Summary*. Washington, D.C., U.S. Dept. of Comm., 20 pp.

NOAA, 1998: *Recommendations from the NOAA Long-term Climate Monitoring Council first meeting.*

Internal NOAA memorandum, March 27, 1998, Asheville, N.C., U.S. Dept. of Comm., 9pp.

NRC (National Research Council), 1998: *Toward a new National Weather Service - Future of the National Weather Service Cooperative Observer Network.* Washington, D.C., National Academy Press, 96pp.

NWS, 1953: *Multiple Address Letter No. 37-53. Climatological Data Substation Network Plan,* Washington, D.C., U.S. Weather Bureau, U.S. Dept. of Com.

WCP-85, 1985: *Guidelines on the quality control of surface climatological data.* World Climate Program Report 85, WMO, 54 pp.

WCP-116, 1986: *Guidelines on the selection of reference climatological stations from existing climatological station networks*. World Climate Program Report 116, WMO/TD-No. 130, 25 pp.

WMO, 1983: *Guide to climatological practices*. WMO-No. 100

WMO, 1992: *Guide to WWW data management*. WMO-No. 788.

Appendix A

Detecting Near-Real Time Daily Maximum and Minimum Temperature Inhomogeneities in the NOAA Cooperative Observer Network

Claude Duchon and Matthew Menne

I. Random Model for Daily Temperature Differences

A. Motivation

We desired to develop a variable that would produce a random series under normal circumstances and a nonrandom series if there were a problem in the observations. If, in addition, when no problems are present, the variable is uncorrelated in time, then simple tests are available for detection of nonrandomness. The variable we chose was the difference in standardized daily temperature departures between nearby station pairs. It is common to refer to the station being evaluated as the *candidate* station and to nearby stations as *neighbors*. The number of acceptable neighbors, *n*, depends on the maximum allowable distance between the candidate and neighbor, if any, and the number of neighbors within this radius that share common observation conditions (e.g. elevation, time of observation). The formula for this variable is

$$D_{Ci,\,Nni} = \frac{(T_{Ci} - \overline{T}_C)}{S_C} - \frac{(T_{Nni} - \overline{T}_{Nn})}{S_{Nn}} \qquad (1)$$

where,

$D_{Ci,Nni}$ = maximum or minimum temperature difference between candidate and neighbor *n* on the i-th day month

T_{Ci} ,T_{Nni} = maximum or minimum temperature at candidate and neighbor *n* on the i-th day

$\overline{T}_C$, $\overline{T}_{Nn}$ = monthly mean maximum or minimum temperatures at candidate and at neighbor *n*

S_C, S_{Nn} = standard deviation of maximum or minimum temperatures at candidate and neighbor *n*

The concept behind this formulation is that if two stations are experiencing the same synoptic regime, the monthly time series of daily differences of departures between two stations should be random and uncorrelated. Local effects produce the randomness. The reason for standardizing the daily departures is to take into account situations where there may be a systematic difference in daily variability. For example, if one of the two comparing stations is under maritime influence it may experience the moderating effect of the nearby water body. While its daily departure from its own monthly mean would be of the same sign as the daily departure of the inland station, the magnitude of the maritime station's departure may be consistently less than at the inland station. This may lead to departure differences that are consistently of one sign (and therefore not random). Standardization helps to alleviate this problem.

B. Test for randomness

A non-parametric test called the *runs test* was selected. Non-parametric tests are useful since no particular underlying probability distribution (e.g., Gaussian) is required. For a random time series with zero autocorrelation the population distribution of runs of like sign has been derived. If there are too few or too many runs the null hypothesis of random differences is rejected. Too few runs might indicate the presence of a trend or a sudden change in the level of temperature at some point during the month being examined. Too many runs might indicate a temperature too high one day, too low the second day, too high the third day, etc. Situations in which there would be too many runs seemed unlikely, so the test was made for too few runs only (a one-sided test). For 30 or 31-day months, the number of runs at the 5% level of significance is 11. Thus, if there are fewer than 11 runs, a flag is raised.

C. Results

The runs test was applied to published COOP temperature data for the north central states for January and July 1997. The potential number of candidate stations was approximately 1200 and candidate stations were paired with their five nearest neighbors. The results of these tests indicated that the number of occurrences of series with sign changes less than or equal to 11 was slightly higher than expected from sampling theory. It was determined that fewer runs than expected can result in a daily departure series as a consequence of the precision of the temperature readings which are recorded to the nearest whole °F. This issue is mitigated to some degree by using standardized departures, but using values recorded to the nearest whole degree inflates the number of runs test flags by at least of few percent. Other runs tests were performed on data from Colorado and North Carolina to determine the effect of mountainous regions on the runs test. It was found that in mountainous areas, local climate effects are sometimes strong enough that daily temperature departures are occasionally of different signs at nearby stations. This can make the detection of errors in mountainous regions more challenging. The effect of mixing observation times and distance on station pairs was also investigated. A candidate station's neighbors must share a similar time of observation to conform to the assumptions using the variable $D_{Ci,Nni}$, namely, that the two stations are tracking one another due to general synoptic influences. Distance between stations was not found to be of concern at least in the eastern two-thirds of the U.S. where station density is high enough for a sufficiently large pool of neighbors can be found within 50 or 60 km.

Many examples of paired difference series that contained less than the expected number of runs were investigated from stations located in north central states and few were found to contain obvious errors. When the level of significance was reduced to 0.4%, corresponding to 7 runs, there were four candidate/neighbor daily departure difference series that had 7 or fewer runs. Each was inspected for errors, but even in these examples none were apparent. The conclusion is that either the runs test is not sufficiently sensitive to detect significant data errors in the daily maximum and minimum temperatures or there are no significant errors.

In the investigation of the runs test, however, it was observed that there were a number of occasions in which the time series of temperature differences between the candidate station and a neighbor and the time series of temperature differences between the candidate station and another neighbor tracked each other quite closely. That is, we compared $D_{Ci,N1i}$ with $D_{Ci,N2i}$. In a large majority of the cases in which there was strong correspondence between the two time series of differences the candidate station had some type of a problem. To quantify the correspondence we calculated the linear correlation coefficient between the monthly time series of $D_{Ci,N1i}$ and $D_{Ci,N2i}$, $D_{Ci,N1i}$ and $D_{Ci,N3i}$, and so on. Calculation of the correlation coefficient between pairs of station differences according to the random model in Equation (1) seems to provide a quite sensitive scheme for error detection in contrast to the insensitivity of the runs test. This scheme is called the *correlation test* and is discussed in more detail below. Examples of the kind of errors that can be detected are discussed and plots of the frequency of occurrence of errors have been provided in the main body of this report (Figures 13a and 13b).

II. Correlation Test

While evaluating the runs test it was noticed that frequently there were a sufficiently large number of sign changes in daily departure difference time series suggesting that the series between a candidate station and its neighbors was random, yet, the daily departure difference series themselves were highly correlated. In certain cases, all of the station paired-difference series combinations are highly correlated. In a sub-sample of 20 of these cases, comprised of stations located in the North Central U.S. using observations from January and July 1997, all but two candidate stations had identifiable problems in the published archive, yet all of them “passed” the runs test. Most of the problems involved a shift in the daily observations of the candidate station with respect to its neighbors for stretches of days during the month. In other words, the candidate station was a partial shifter. Four correlation coefficients were then routinely calculated that paired the candidate station with its five nearest neighbors. The correlation test is based on the value of Pearson’s correlation coefficient between the following daily departure difference series: $D_{Ci,N1i}$ and $D_{Ci,N2i}$, $D_{Ci,N1i}$ and $D_{Ci,N3i}$, $D_{Ci,N1i}$ and $D_{Ci,N4i}$, $D_{Ci,N1i}$ and $D_{Ci,N5i}$.

In a sub-sample of 15 examples where 3 of the 4 station paired-difference series were highly correlated (correlation coefficients greater than 0.75), while the fourth pair was lower, there were fewer identifiable problems in the data. In examples where all four paired-difference correlation coefficients were very high, above 0.8, definite problems in the data were evident and easily identifiable. So, it appears that correlation among paired-differences is a more sensitive test for data errors than the runs test. The summary statistics displayed in Figure 13 of Section 2.3.7 were compiled by flagging stations where each of the four calculated correlation coefficients was greater than 0.7.

The reason that high correlation among paired differences can result from observation shifting is illustrated in the following example. COOP Station 115901, Mount Carroll, IL, contained no high correlation coefficients among station daily departure differences for maximum temperatures during the month of July 1997 and each of the departure series had more than 11 sign changes (see Table A1). These statistical measures, coupled with visual inspection, all suggest that the original data are valid at this station (something about the operational QC pass?). When the maximum temperature recorded on July 21 was shifted back (i.e. attributed to) July 20, the correlation among departure difference series increased substantially, yet the number of sign changes from the runs test fell below the threshold for a random series at the 5% probability in only one case. The number of runs and the values of the correlation coefficients for a shift that occurs for a two-day duration, July 20-21, are also presented in Table A1. In this case, the number of sign changes in all cases was within the range that would suggest a random time series. The daily departure difference series, however, were all highly correlated.

The effectiveness of the correlation test to detect observation errors can depend on where in a time series shifts occur. To illustrate, a three-day stretch at Mount Carroll was shifted back one day during a period when daily maximum temperatures were trending upward (July 12-14). In this case, shifting the maximum temperatures for three days had little impact on any method of error detection. Indeed, even upon visual inspection, it is essentially impossible to detect that a shift has occurred.

Figure A1 (top) shows a plot of the published daily maximum temperatures at Mount Carroll; the two-day shift is indicated by the bold dashed line and the daily departure differences that result from this partial shift are plotted in the bottom figure. Figure A2 is scatter plot of two of the daily departure difference series ($D_{Ci,N1i}$ vs. $D_{Ci,N2i}$) and illustrates how high correlation coefficients between daily departure differences result from data errors. Two of the observations appear to be outliers; these are the paired differences for July 20 and July 21. Without the strong influence of these two points, the correlation coefficient between these two daily departure difference series would be 0.462 rather than

Table A1. Number of sign changes in daily departure difference series and correlation coefficients between paired departure difference time series during July 1997 for various imposed partial shifts in maximum temperature observations at the Mount Carroll, IL COOP station

Example	Number of sign changes (runs) in daily departure difference series between candidate, C, and five nearest neighbors, N1 – N5					Correlation coefficients between four daily departure difference series			
	$D_{Ci,N1i}$	$D_{Ci,N2I}$	$D_{Ci,N3i}$	$D_{Ci,N4i}$	$D_{Ci,N5i}$	$D_{Ci,N1i}$ vs. $D_{Ci,N2i}$	$D_{Ci,N1i}$ vs. $D_{Ci,N3i}$	$D_{Ci,N1I}$ vs. $D_{Ci,N4I}$	$D_{Ci,N1i}$ Vs. $D_{Ci,N5i}$
Original (unaltered) Observations	15	14	14	15	17	0.440	0.262	-0.308	0.270
One-day shift occurring July 20	15	9	14	14	18	0.855	0.706	0.578	0.753
Two-day shift occurring July 20-21	13	16	14	15	18	0.907	0.805	0.754	0.846
Three-day shift occurring July 12-14	15	11	12	14	18	0.532	0.371	-0.026	0.336

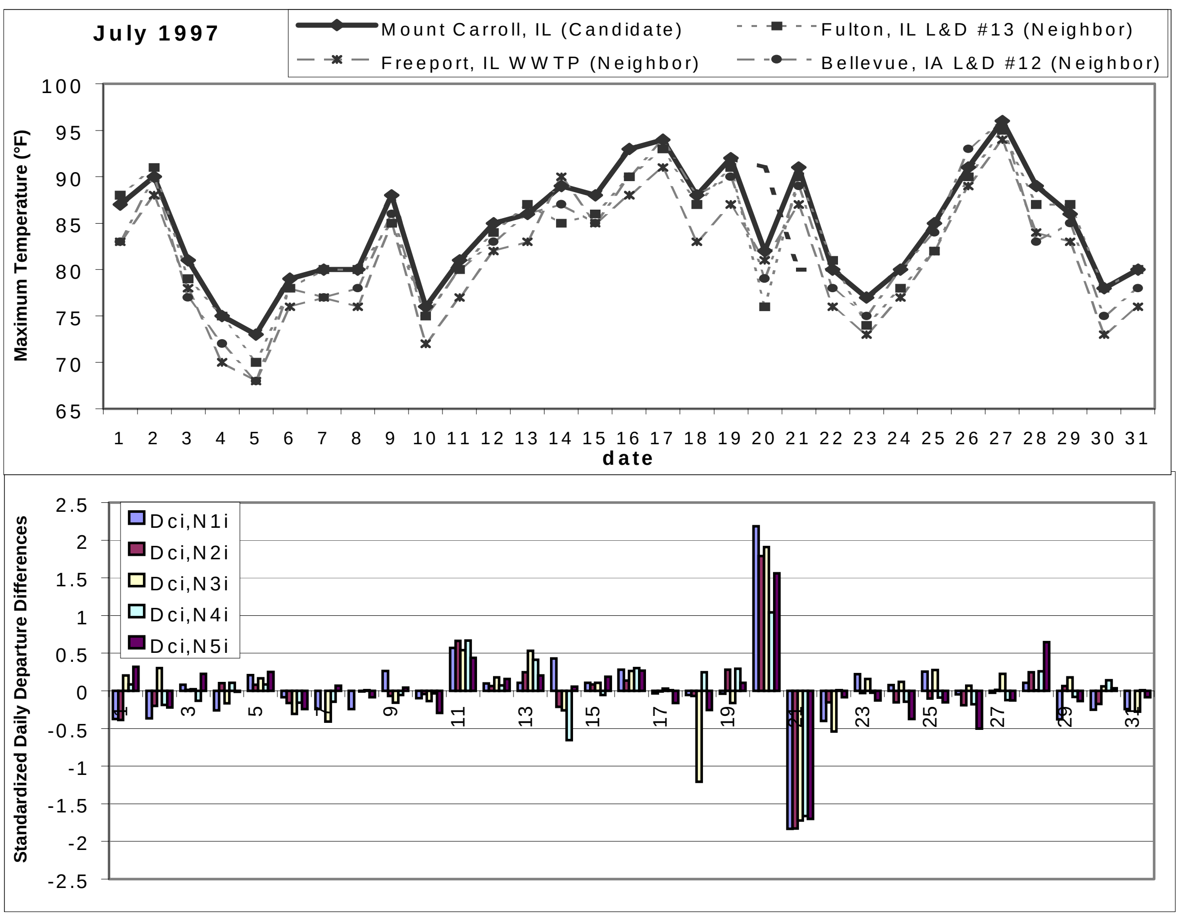


Figure A1. Daily maximum temperatures at candidate Mount Carroll and three neighbors (top) and standardized daily departure differences (bottom) for candidate and neighbors for a two-day partial shift. The bold dashed line in top figure depicts the two-day shift.

0.910. It is thought that the nature of the departure differences can be used in an automated procedure for error detection since they can be used to differentiate among the types of detectable error. It should be also noted that with a sample size of around 30 observations (1 month) and the potential for serial correlation when errors are present, the errors associated with estimating the correlation between series are large. The correlation coefficient is used here, however, simply to identify problems in the data.

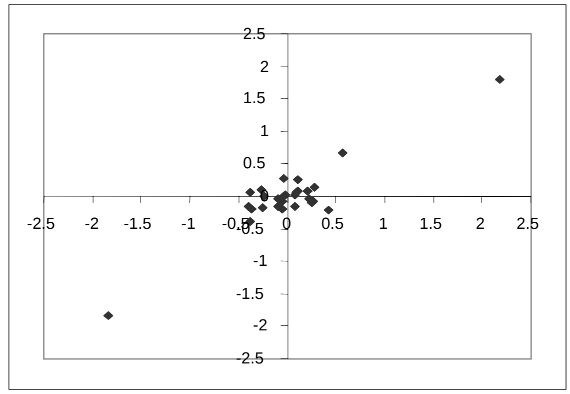


Figure A2. Scatter plot of daily departure differences ($D_{Ci,N1i}$ and $D_{Ci, N2i}$) for a two-day partial shift at Mount Carroll, IL.

Two examples of errors that were detected using the correlation test are shown from the archived data from the sample of stations located in the north central states. The first example is presented in Figure 3A and shows the candidate station Chicago O'Hare and its neighbors. The maximum temperature observations for a stretch of days beginning around January 22, 1997 are out of range with respect to its neighbors. Chicago O'Hare is both an NWS ASOS station and a COOP station. Records from ASOS stations that are also COOP observers follow a different quality control path and this path will change in the future. Consequently, this error might be caught in the future using current quality control methods. Nevertheless, this example illustrates how

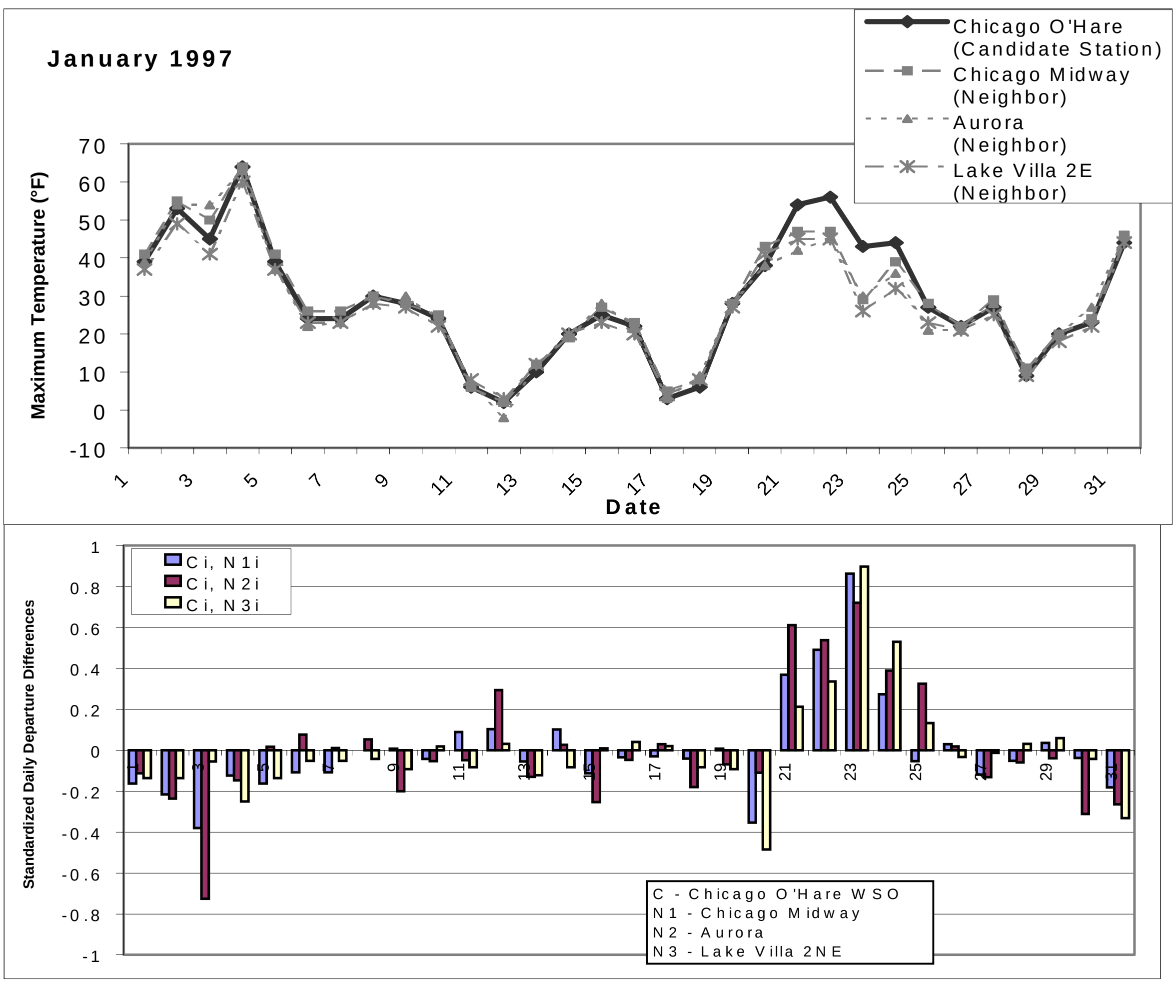


Figure A3. Daily maximum temperatures at candidate station Chicago O'Hare and nearby neighbors (top) and standardized departure differences between candidate and neighbors (bottom).

drifts in a temperature sensor may be identified in near-real time using the correlation test and is used as an example of how the correlation test can identify inhomogeneities that affect the monthly mean temperature.

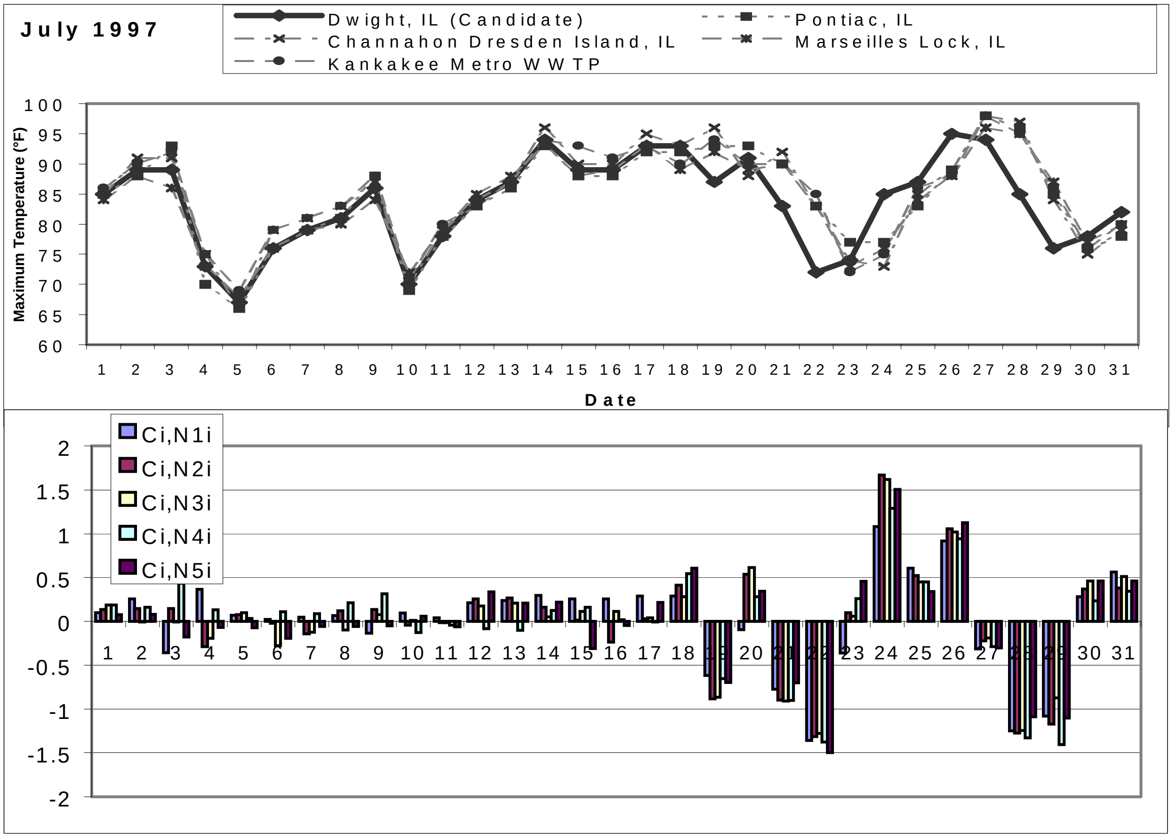


Figure A4. Daily maximum temperatures at candidate station Dwight, IL and nearby neighbors (top) and standardized departure differences between candidate and neighbors (bottom).

The example at Dwight, IL shown in Figure 4A is a partial shifter. Observations at the candidate station were shifted starting on July 18 and continued into the first few days of August after which the observer got back on track. Figure A5 is a scatter plot between two departure difference series. Notice that a strong linear correlation is evident when a series of days are shifted.

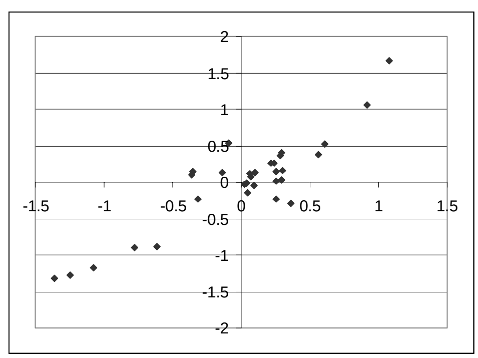


Figure A5. Scatter plot of daily departure differences ($D_{Ci, N1i}$ and $D_{Ci, N2i}$) at Dwight, IL.

III. Detection of Trends

A. Theory

One approach to detecting a linear trend in a stationary time series is to apply linear regression theory. In this approach the null hypothesis is made that the slope β of the population regression line is zero, that is, H_o: β= 0. If the departure of the observed trend from 0 is significant, the null hypothesis of no trend is rejected.

The linear first-order regression model is

$$Y = \alpha + \beta \cdot X + \varepsilon \tag{2}$$

where Y is the predictand, X the predictor, α is the Y-axis intercept, and ε is a random departure of any individual value of Y from the regression line. In practice, we don't know ε so that α and β are estimated using the formula

$$\hat{Y} = a + b \cdot X \tag{3}$$

where $\hat{Y}$ is the predicted value of X when a and b are determined. The minimum least-squares estimation of b is given by

$$b = \frac{\sum x \cdot y}{\sum x^2} \tag{4}$$

in which $x = X - \overline{X}$ and $y = Y - \overline{Y}$. The null hypothesis H_o: = 0 is tested using the t-statistic

$$t = \frac{b - \beta}{S_b} = \frac{b - 0}{S_b}\,. \tag{5}$$

If the magnitude of the t-statistic exceeds the critical value t_c associated with a chosen confidence interval, often 95% or 99%, the null hypothesis is rejected, otherwise it is not.

Estimated trend b is distributed with variance

$$S_b^2 = \frac{S_{x \cdot y}^2}{\sum x^2} \tag{6}$$

The quantity $S_{x \cdot y}^2$ in (6) is the variance of the predicted minus observed differences, i.e.,

$$S_{x \cdot y}^2 = \frac{\sum (\hat{y} - y)^2}{(n - 2)} \tag{7}$$

in which $\hat{y} = a + b \cdot x$ and n is the number of data.

The t-test is valid if y is a Gaussian-distributed random variable and the residuals are uncorrelated.

B. Application

In applying the first-order linear regression model to coop data, the variable y is the same as the variable $D_{Ci,Nni}$ in Equation (1) except that the daily departures are not standardized. It is assumed that the differences are approximately normally distributed. The variable x is the day of a month (i.e., 1 through 31 for one month of data, 1 through 62 for 2 consecutive months of data, etc.; thus n = 31, n = 62, respectively, etc.). Now let us select a two-tail 5% level of significance (95% confidence interval) so that there is 2-1/2% at either end of the t-distribution. Under this condition and assuming the residuals $(y - \hat{y})$ are uncorrelated, the critical value of the t-statistic from (5) is

$$t_c = \frac{|b_{min}|}{S_b} = 2.045 . \tag{8}$$

From (6) and (7) we see that S_{BA} is dependent on the number of days n and variance of the residuals, i.e., the observed differences of departures minus the fitted straight-line. Thus we can calculate the minimum detectable linear trend in the differences of the departures according to

$$|b_{min}|[\text{F deg/month}] = S_b \times 2.045 . \tag{9}$$

Figure A6 is a plot of $|b_{min}|$ for 2 levels of significance versus the number of months over which the trend is occurring for three variances of the residuals from Equation (7). When the coefficients a and b are zero, $S^2_{x \cdot y}$ is essentially the variance of a departure difference time series.

Interpreted in thus light, Figure A6 shows that as the variability in the differences between two stations increases for a given period, the minimum detectable trend increases. Trends in the neighborhood of a few degrees F per month will require a few months to detect. Trends around a few degrees per year will require many months to detect.

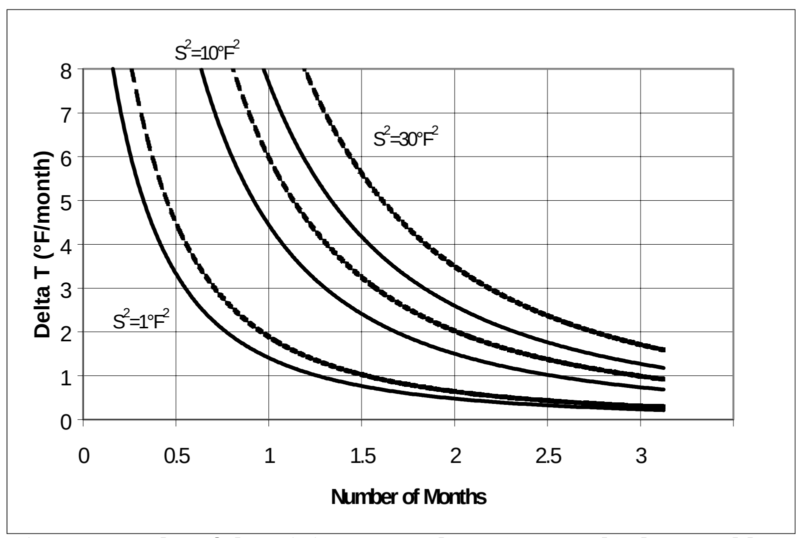
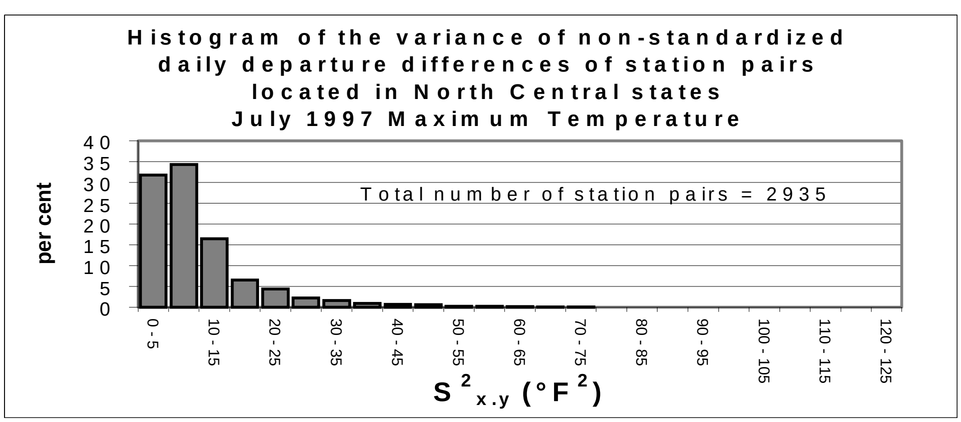


Figure A6. Plot of the minimum trend, ΔT (°F/month) that would just reject the null hypothesis of no trend at the 5% (solid) and 1% (dashed) significance level for typical ranges in the variance, S^2, of the observed differences of departures minus the straight-line fit *versus* the length of the time series in months.

Figure A7 shows histograms of the observed variances of the departure difference time series for stations in the north central states for 1997 January and July maximum temperatures. The variances are computed with coefficients **a** and **b** equal to zero as discussed above. The range encompass ctedly large variances these large variances v



C. Interpretation

It is important to recognize that the theory and application described above assume that the trend magnitudes follow a t-distribution (or normal distribution if n is large). The rejection criterion is applicable to selecting, at random, a pair of stations and performing the test on the one pair. When there are many pairs of stations being compared, say, 100, one would expect, on the average, 5 pairs to be rejected. There will have to be substantially more than 5 rejected to raise concern. Alternatively, one can calculate a new set of confidence limits such that there is only a 5% chance that the trends in any one or more of the 100 pairs of differences would be rejected. While this approach is statistically sound, it may be easier and more fruitful to develop a distribution of observed trends among paired stations over 1 month, 2 months, 3 months, etc., until a sufficiently narrow distribution of trends is observed such one can make a judgment, based on physical grounds, as opposed to statistical grounds, whether there is evidence for a trend in any pair of stations. A first step in this direction would be to explain the very large variances in Figure A7.

D. Future Work

As mentioned above, continuing work will proceed with the development of an observed distribution of trends in order to determine more carefully what magnitudes of trends over short time scales (one to a few months) may be large enough to be of concern, if any. This step will be followed by an assessment of the sensitivity of the correlation test in detecting true inhomogeneities. It has been noted in work not reported here that the correlation between daily departure difference series increases when linear trends or step changes are imposed on observed data. It has also been demonstrated that the correlation test can identify trends or step changes under some circumstances and its sensitivity will be evaluated in light of the observed variability and distribution of trends in daily maximum and minimum temperatures.